6. The apparatus of claim 1, wherein each of said at least one audio command and said speech recognition data comprises first and second command word portions separated by said pause portion and further comprising a second pause portion having one syllable in duration before said first command word portion and a third pause portion having one syllable in duration after said second command word portion.

27

7. The apparatus of claim 1, wherein the speech recognition means further including a microcontroller with a fixed-point embedded microprocessor, the microprocessor is chosen from the group of 8-bit and 16-bit micro controller unit_microprocessors.

8. A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:

recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;

receiving at least one audio command from a user, the at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length;

comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;

generating at least one control signal based on said comparison;

controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command;

analyzing the pause portion of the received audio command for spectral content; and

97

preventing operation of the electrical device when the spectral content is dynamic.

9. The method of claim 8, wherein the step of recording speech recognition data includes recording the voice of a user while the user utters said at least one audio command.

12. The method of claim 8, and further comprising:

23

ascertaining a first energy content for the command word portion of the received audio command;

ascertaining a second energy content for the received background noise data;

a3

comparing the first and second energy contents and generating an energy comparison value; and

preventing the generation of said at least one control signal when said energy comparison value is below a predetermined level.

7

16. The apparatus of claim 8, wherein each of said at least one audio command and said speech recognition data comprises at least first and second command word portions separated by said pause portion.

a4

12

17. The apparatus of claim 14, and further comprising a second pause portion having one syllable in duration before said first command word portion and a third pause portion having one syllable in duration after a second command word portion.

---

27. A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:

a5

recording speech recognition data having a command word portion and a pause portion, each of the speech recognition data portions being at least one syllable in length;

receiving at least one audio command from a user, the at least one audio command having first and second command word portions and a first, second and third pause portions, each of the audio command portions being at least one syllable in length,

said second pause portion having one syllable in duration before said first command word portion and said third pause portion having one syllable in duration after said second command word portion;

comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech recognition data;

generating at least one control signal based on said comparison;

controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command.

In the Abstract of the Disclosure

Please substitute the Abstract presented hereinbelow for the Abstract currently of record.

ABSTRACT OF THE DISCLOSURE

A voice activated apparatus consists of a receiving arrangement for receiving an audio command having command word and pause portions. A speech recognition data having command word and pause portions. Speech recognition arrangement generating a control signal based on comparing of the command word portion and the pause portion of the audio command with the command word portion and the pause portion of the speech recognition data. A power control arrangement controls power delivered to an electrical device in response to the control signal generated by the speech recognition arrangement in response to the audio command.

II Remarks

The undersigned wishes to thank Examiner Daniel A. Nolan for the courtesy extended to him during telephone conferences conducted during preparation of the Response.

As to the required correction of the drawings, amended FIGs. 2 and 8 are enclosed for the approval by the Examiner.

As required by the Examiner, a Substitute Specification containing the revised Background of the Invention, Summary Of The Invention, Detailed Description of the Preferred Embodiment and Abstract of the Disclosure accompany the Response. It should be noted, however, that <u>the Substitute Specification includes the originally filed set of claims</u>. <u>The amended set of claims is presented in this Response</u>.

As to the Examiner's remarks presented in paragraph 7 of the Office Action, the following amendments are made to the specification. The term "speech recognition" was substituted throughout the specification for the term "voice recognition". The plural word "Figures" was introduced in the first paragraph on former page 23 of the Application. The Summary Of

The Invention has been amended to satisfy the requirements of MPEP, Paragraph 608. 01(b). The term "microcontroller unit" has been inserted next to the acronym "MCU" on the former page 12 of the Application.

Furthermore, the specification has been checked to the extent necessary to determine the presence of minor errors. All the above-discussed amendments are reflected in the attached Substitute Specification.

In Paragraph 9 of the Office Action the Examiner indicates that the specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. Specifically, the Examiner indicates that the disclosure that a pause may be substituted for a syllable (see page 6) will not support either a claim where the pause is measured in terms of syllables, nor a potential claim of a pause consisting of multiple syllables (see claims 6 and 17). A more appropriate unit to designate measurements and duration, or for representation of multiple pauses is required by the Examiner.

Applicants respectfully disagree with the above-discussed position of the Examiner. A syllable is the perfect measure of a pause. In this

respect, it is a well known fact that a pause between quickly spoken words is shorter than a pause between slowly spoken words.

There is clear evidence that poets and writers often substitute pauses for syllables. This occurs, for example, when a poet or writer needs to insert pauses, so the number of syllables for each line of poetry is consistent. Exhibits A and B demonstrate this usage. It is clearly illustrated in the attached Exhibit A that (1) a syllable is a commonly accepted measure of the duration of pause, and (2) in terms of measuring the length of a pause, the syllable is a clear alternative for using the units of time (such as minutes) for the same purpose. The attached Exhibit B further demonstrates that it is known for a pause to be measured in terms of syllable or that a pause can be measured by multiple syllables.

This technique has been fully explained and further developed by the present application. In this respect, please refer, for example, to the paragraph bridging pages 19 and 20 of the originally filed case. For example, it is specified there that "The duration of the pause model 164 between each command word may vary depending on the particular speaking style of the user (ers) but should be at least one syllable (about 200

msec.) in length." It is further indicated there that to increase a command detection accuracy, a pause may be added before and/or after each command word. "Thus, instead of a two-syllable command for "lightson", a three-syllable command "lights<pause>on" increases the FOM, while a five-syllable command "<pause>lights<pause>on<pause>" greatly increases the FOM without increasing the number of words in the voice command."

The language of the former page 6 Application (see the attached Substitute Specification) has been slightly amended to clarify this feature of the invention. It is respectfully considered that no new matter is introduced into the Application by this amendment.

Applicants have noted that claims 2-5, 10-15 and 17 are objected to as being dependent upon the rejected base claim but would be allowable if re-written in the form suggested by the Examiner in the Outstanding Office Action.

Claim 1 has been amended to include the limitations of the allowable claim 2 and therefore is allowable. Claims 3, 4, 5, 6 and 7 are directly or indirectly dependent upon allowable claim 1, do not contain independently

patentable subject matter and therefore should be allowable. Independent claim 8 has been amended to incorporate the limitations of the allowable claim 10 and therefore is also allowable. Claims 9 and 11-17 are directly or indirectly dependent upon allowable claim 8, do not include independently patentable subject matter, and also should be allowable. New claim 27 represents allowable claim 17 re-written in independent form and also should be allowable.

Claim 17 has been amended to overcome the Examiner's rejection under 35 USC 112, 2nd paragraph, as being indefinite. The amended version of claim 17 satisfies the requirements of 35 USC 112 and particularly point out and distinctly claim the subject matter of the invention.

The claims 2, 10 and the claims 18-26 have been canceled without prejudice. Thus, the claims amendment presented in the Response should place all claims remaining in the Application in condition for allowance.

In the Outstanding Office Action claims 1, 6, 8, 9 and 16 were rejected under 35 USC 103(a) as being unpatentable in view of U.S. Patent

6,208,971 to Belegarda, et al. in view of U.S. Patent 5,086,385 to Launey, et al. Claim 7 has been rejected under 35 USC 103(a) as being obvious over the Belegarda, et al reference in view of the Launey, et al. reference and further in view of U.S. Patent 5,890,121 to Borcherding. Applicant respectfully considers that this rejection of the Examiner is moot in view of the amendment of the claims presented in the Response.

Withdrawal of the Examiner's rejection and allowance of all claims currently of record in the Application are respectfully requested in view of the Amendment and Remarks presented in the Response.

Applicant has made the best faith effort to place the referenced Application in condition for allowance. However, if any issue raised by the Patent & Trademark Office has inadvertently been left unanswered, the Examiner is authorized to call the undersigned at the telephone number indicated hereinbelow.

Applicant respectfully petitions for three months extension of time for reply. A separate Petition and a check in the amount of $ 460.00 are enclosed.

III Version Of Claims And Paragraphs Of Specification With Markings To Show The Changes Made

Amendment of Claims

1. (amended) An apparatus for voice activated control of an electrical device, the apparatus comprising:

receiving means for receiving at least one audio command generated by a user, [the] at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length;

speech [voice] recognition data having a command word portion and a pause portion, each of the speech [voice] recognition data portions being at least one syllable in length;

speech [voice] recognition means including a Hidden Markov Model for comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech [voice] recognition data,

said speech [voice] recognition means generating at least one control signal based on said comparison, said speech recognition means prevents operation of the electrical device when the spectral content is dynamic;

means for analyzing the pause portion of the received audio command for spectral content; and

power control means for controlling power delivered to an electrical device, said power control means being responsive to said at least one control signal generated by said speech [voice] recognition means for operating the electrical device in response to said at least one audio command generated by the user.

3. (amended) The apparatus of claim 1, wherein said receiving means receives background noise data in conjunction with said audio command, and further comprising means for generating a command word score and a background noise score based on the comparison of the received audio command to the speech [voice] recognition data and the background noise data, respectively, said speech [voice] recognition means generating said at least one control signal when said command word score exceeds said background noise score.

4. (amended) The apparatus of claim 3, and further comprising:

means for analyzing the command word portion of the received audio command and the background noise data for energy content; and

means for comparing the energy content of the command word portion to the energy content of the background noise data and generating a corresponding energy comparison value;

wherein said <u>speech</u> [voice] recognition means prevents the generation of said at least one control signal when said energy comparison value is below a predetermined level.

5. (amended) The apparatus of claim 1, wherein said receiving means receives background noise data in conjunction with said audio command, and further comprising:

means for analyzing the command word portion of the receive audio command and the background noise data for energy content; and

means for comparing the energy content of the command word portion to the energy content of the background noise data and generating a corresponding energy comparison value;

wherein said speech [voice] recognition means prevents the generation of said at least one control signal when said energy comparison value is below a predetermined level.

6. (amended) The apparatus of claim 1, wherein each of said at least one audio command and said speech [voice] recognition data comprises [at least] first and second command word portions separated by said pause portion and further comprising a second pause portion having [at least] one syllable in duration before said first command word portion and a third pause portion having [at least] one syllable in duration after said second command word portion.

7. (amended) The apparatus of claim 1, wherein the speech [voice] recognition means further including a microcontroller with a fixed-point embedded microprocessor, the microprocessor is chosen from the group of 8-bit and 16-bit [MCU] micro controller unit microprocessors.

8. (amended) A method of activating an electrical device through at least one audio command from a user, the method comprising the steps of:

recording speech [voice] recognition data having a command word portion and a pause portion, each of the speech [voice] recognition data portions being at least one syllable in length;

receiving at least one audio command from a user, the at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length;

comparing said command word portion and said pause portion of said at least one received audio command with said command word portion and said pause portion, respectively, of said speech [voice] recognition data;

generating at least one control signal based on said comparison; [and]

controlling power delivered to an electrical device in response to said at least one control signal for operating the electrical device in response to said at least one received audio command;

analyzing the pause portion of the received audio command for spectral content; and

preventing operation of the electrical device when the spectral content is dynamic.

9. (amended) The method of claim 8, wherein the step of recording speech [voice] recognition data includes recording the voice of a user while the user utters said at least one audio command.

12. (amended) The method of claim 8 [10], and further comprising:

ascertaining a first energy content for the command word portion of the received audio command;

ascertaining a second energy content for the received background noise data;

comparing the first and second energy contents and generating an energy comparison value; and

preventing the generation of said at least one control signal when said energy comparison value is below a predetermined level.

16. (amended) The apparatus of claim 8, wherein each of said at least one audio command and said speech [voice] recognition data comprises at least first and second command word portions separated by said pause portion.

17. (amended) The apparatus of claim 14, and further comprising a second pause portion having [at least] one syllable in duration before said first command word portion and a third pause portion having [at least] one syllable in duration after [said] a second command word portion.

Respectfully submitted,
SILBER & FRIDMAN

Lawrence G. Fridman
Registration No. 31,615
Attorney for Applicant

66 Mount Prospect Avenue
Clifton, New Jersey 07013-1918
Telephone (973) 779-2580
Fax (973) 779-4473
RESP01_618A

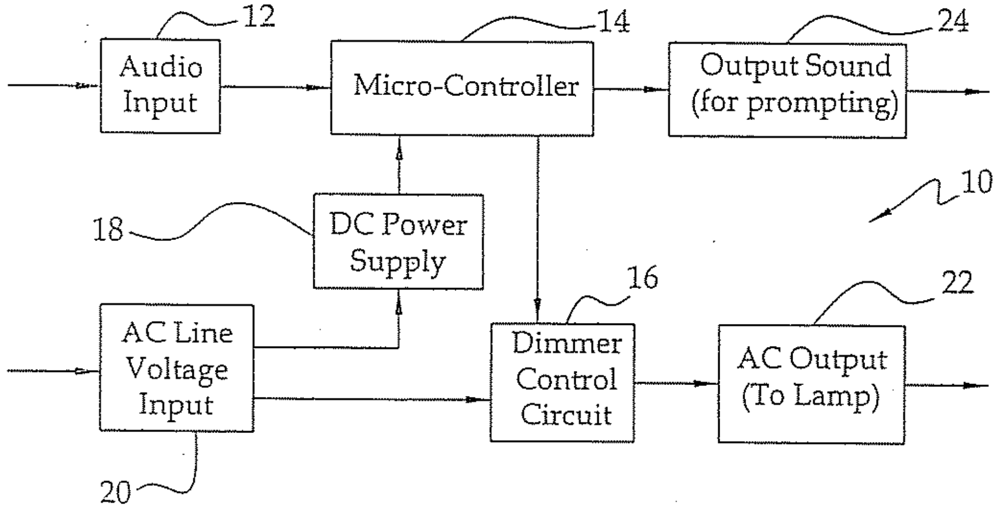


FIG. 1

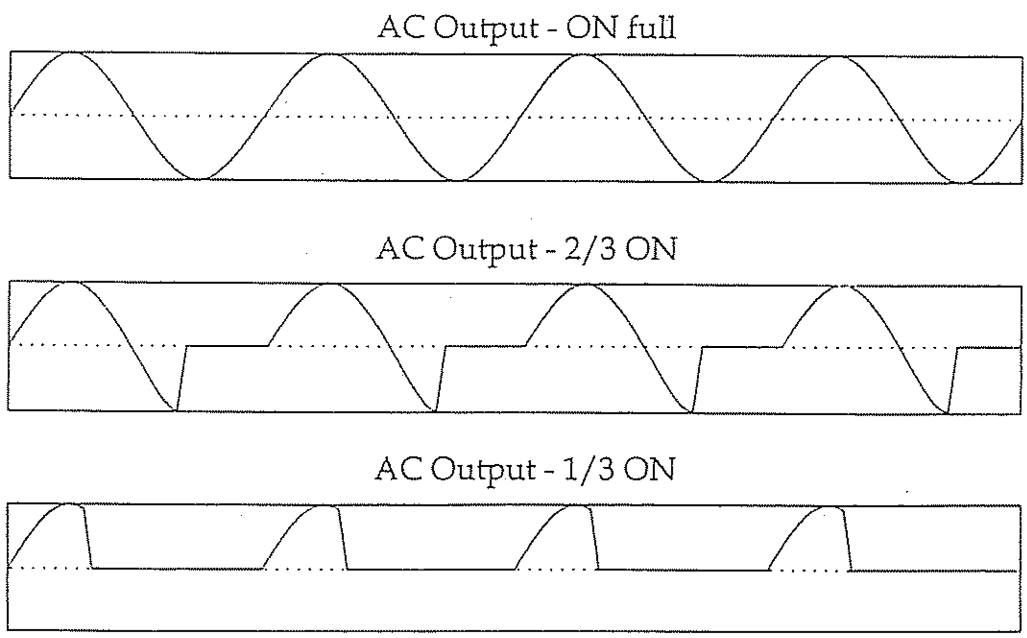
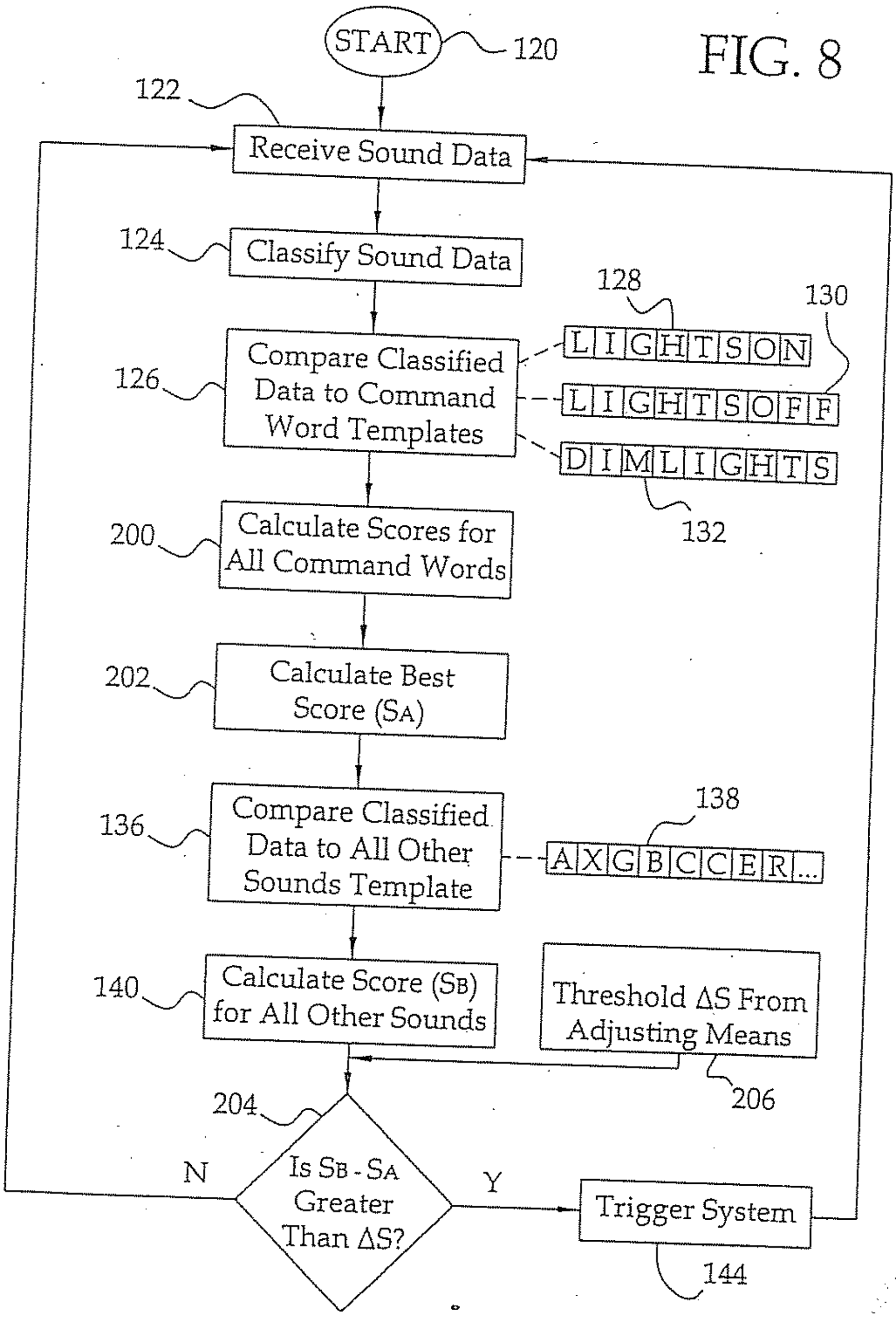


FIG. 2 PRIOR ART

FIG. 8


START
120
122
Receive Sound Data
124
Classify Sound Data
128
130
126
Compare Classified Data to Command Word Templates
LIGHTSON
LIGHTSOFF
DIMLIGHTS
132
200
Calculate Scores for All Command Words
202
Calculate Best Score (SA)
138
136
Compare Classified Data to All Other Sounds Template
AXGBCCER...
140
Calculate Score (SB) for All Other Sounds
Threshold ΔS From Adjusting Means
206
204
N
Is SB - SA Greater Than ΔS?
Y
Trigger System
144

The salon.com Readers' Guide to Contemporary Authors

mediacircus under the covers

S A L O N

I can't get arrested in this town!

A TALE OF THE . . . thin blue . . . ROPE LINE.

BY JAMES PONIEWOZIK

> "An ongoing New York protest against the police shooting of an unarmed street vendor got a dose of Hollywood support this week. Activist Susan Sarandon and 218 others were arrested Thursday ..." -- "Names & Faces," Washington Post, March 27

Ronnie, hi, it's -- sorry honey, can you hold?

[Eight-minute pause.]

Hi, Ronnie, you're a sweetheart, that was my homeopath, he's impossible to actually get on the phone. So I called because the talk we had last week? About this is a transitional point in my career and we need to raise my profile and get me into quality projects? Well, I'm doing Pilates with the cable news on -- it helps me clear my mind and get centered -- and guess who I see? Susan Sarandon, on the television, prime time, getting arrested. Looking absolutely fucking drop-dead gorgeous, I could rip her throat out. Have you heard about this Amadou Diablo man in New York? Well, apparently the thing with this Diablo --

[Three-syllable pause.]

You see, Ronnie, this is exactly what I was talking about. The second-guessing and the corrections and the nit-picking. Do I have to remind you who could have jumped over to Ovitz when he came sniffing

TABLE TALK

So you've seen the first episode of Matt Groening's "Futurama"? Wha'dya think? Join the new discussion in Table Talk

Extra! Extra! Find the media books you want to read at barnesandnoble.com

Search by: Keyword Go

RECENTLY

$400,000 misunderstanding
By Susan Lehman
So maybe Vanity Fair writers don't actually make 400 grand. Who needs

Exhibit 1

PH 101

DRAGON TAILS MESSAGE BOARD!

General forum
Roleplaying forum
Art forum
Polished Art forum
Literary forum

Message Board Rules - *you must read these before you post on this board*

Dragon Tails message boards
> Literary forum
> Stealth Obsidian("The End")

New Topic Reply

Page 1 2 3 4 5 6 7 8 9 10

<< Prev Topic | Next Topic >>

Author	Comment
Sapien X99 **Registered User** (1/25/01 7:01:02 am) Reply 	**Stealth Obsidian("The End")** *" Come Vengeance "* Innocence Has lost her head; Justice, Where have you fled? Come Vengeance, Out of the night: Hunters Sure of their might, For Blood, Soon to be shed With blade With axe and bow, For reasons You cannot know-- Perhaps, The price was right? *Edited by: Sapien X99 at: 5/1/02 8:34:48 am*
Moonstone Spider **Registered User** (1/25/01 9:34:20 am) Reply 	**Re: Stealth Obsidian** Good feel and a nice emotive quality, but I can't see what the title has to do with the content. Also you may want to reword and rephrase certain lines as your rythm is screwed in a couple of places. Some lines only have 5 syllables and some have 6 or seven. I like it, though.
Sapien X99 **Registered User** (1/25/01 2:33:09 pm) Reply	**Yah, thanks** Stealth Obsidian is the name of a group of bounty hunters in one of my stories. I guess you have no way of knowing that, though. I thought it would be a good title for my literary thread. Thanks for the feedback. I tried to have six syllables per line, but I was able to add a lot of clarity by making some lines seven syllables. The part that comes before the comma is

Exhibit 2

supposed to be like an interjection, or a brief (2-3 syllable) pause in the poem that states the theme of the line. Thus I was much stricter in forcing the part to the right of the comma to be four syllables exactly. Still, I am not totally happy with it. If you have any other suggestions I am glad to hear 'em. thanks

Sapien X99
Registered User
(2/4/01 3:12:50 am)
Reply

Rock my Brain

" Rock my brain "

Rock my brain, leave my brain.
Why can't I just be sane--
learn to refrain again?

Caught in mud, drawing blood,
Craving death, hating good,
Life too late understood.

{ author's note- this was supposed to sound a little disturbing. It's just a poem, though. so don't worry about me.. too much }

Edited by: Sapien X99 at: 2/4/01 3:28:01 am

Tailsteak
Registered User
(2/4/01 3:55:01 am)
Reply

Disturbing

Well, it's certainly evocative.

Saint by Grace, Worldsmith by Nature, Freak by Choice.

Sapien X99
Registered User
(2/5/01 7:21:18 am)
Reply

Destruction Haiku

" Destruction Haiku "

I saw destruction.
The only thing left standing
was a single cross--

The igneous flames
Had revealed the gold beneath,
The names engraved there...

Edited by: Sapien X99 at: 2/5/01 7:22:43 am

Sapien X99
Registered User
(2/5/01 7:28:45 am)
Reply

Re: Destruction Haiku

{ looks up *evocative* }
thanks, Tailsteak!

Tailsteak
Registered User
(2/6/01 11:38:25 am)
Reply

Haiku

Hmm... this new art form closely paralells "Spamku". I once saw an entire web site devoted to spamku.... I can't remember the addy, and I don't even know if it's still up...

PH 103

File No.: F9018A

UNITED STATES PATENT AND TRADEMARK OFFICE

In re New Application of: Dr. Igor Zlokarnik, et al.

U.S. Application Serial No. : 09/443,957

Filed: November 19, 1999

OIPE JUN 1 1 2002 PATENT & TRADEMARK OFFICE

Group Art Unit: 2641

Examiner: Daniel A. Nolan

For: Voice Activated Control For Electrical Device

RECEIVED JUN 2 0 2002 Technology Center 2600

Assistant Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

TRANSMITTAL OF THE SUBSTITUTE SPECIFICATION

Please find enclosed the Substitute Specification required by the Examiner. The Substitute Specification accompanies the Response to the Office Action mailed December 3, 2001.

MAILING CERTIFICATE

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to Assistant Commissioner for Patents, Washington, D.C. 20231 on May 31, 2002.

Name of person signing: Lawrence G. Fridman ________________
Signature

The undersigned certifies that there is no new matter has been introduced into the Application by the Substitute Specification.

Entering of the Substitute Specification into the Application and considering thereof by the Examiner are respectfully requested.

Respectfully submitted,

SILBER & FRIDMAN

Lawrence G. Fridman
Registration No. 31,615
Attorney for Applicant

66 Mount Prospect Avenue
Clifton, New Jersey 07013-1918
Telephone (973) 779-2580
Fax (973) 779-4473
RESP02_618A

SUBSTITUTE SPECIFICATION

F9618 A

OIPE
JUN 1 1 2002 JC24
TRADEMARK OFFICE

VOICE-ACTIVATED CONTROL FOR ELECTRICAL DEVICES

FIELD OF THE INVENTION

This invention relates to the field of speech recognition and, more particularly, to utilizing human speech for controlling voltage supplied to electrical devices, such as lights, lighting fixtures, electrical outlets, volume, or any other electrical device.

RECEIVED
JUN 2 0 2002
Technology Center 2600

BACKGROUND OF THE INVENTION

The ability to detect human speech and recognize phonemes has been the subject of a great deal of research and analysis. Human speech contains both voiced and unvoiced sounds. Voiced speech contains a set of predominant frequency components known as formant frequencies which are often used to identify a distinct sound.

Recent advances in speech recognition technology have enabled speech recognition systems to migrate from the laboratory to many services and products. Emerging markets for speech recognition systems are appliances that can be remotely controlled by voice commands. With the highest degree of consumer convenience in mind, these appliances should ideally always be actively listening for the voice commands (also called keywords) as opposed to having only a brief

recognition window. It is known that analog audio input from a microphone can be digitized and processed by a micro-controller, micro-processor, micro-computer or other similar devices capable of computation. A speech recognition algorithm can be applied continuously to the digitized speech in an attempt to identify or match a speech command. Once the desired command has been found, circuitry which controls the amount of current delivered to a lighting fixture or other electrical device can be regulated in the manner appropriate for the command which has been detected.

One problem in speech recognition is to verify the occurrence of keywords in an unknown speech utterance. The main difficulty arises from the fact that the recognizer must spot a keyword embedded in other speech or sounds ("wordspotting") while at the same time reject speech that does not include any of the valid keywords. Filler models are employed to act as a sink for out-of-vocabulary speech events and background sounds.

The performance measure for wordspotters is the Figure of Merit (FOM), which is the average keyword detection rate over the range of 1-10 false alarms per keyword per hour. The FOM increases with the number of syllables contained in a keyword (e.g. Wilcox, L.D. and bush, M.A.: "Training and seach algorithms for an interactive wordspotting system" Proc. Of ICASSP, Vol. II, pp 97-100, 1992) because more information is available for decision making. While

using longer voice commands provides an easy way of boosting the performance of wordspotters, it is more convenient for users to memorize and say short commands. A speech recognition system's susceptibility to a mistaken recognition, i.e. a false alarm, generally decreases with the length of the command word. A longer voice command makes it more difficult for a user to remember the voice command vocabulary, which may have many individual words that must be spoken in a particular sequence.

Some speech recognition systems require the speaker to pause between words, which is known as "discrete dictation." The intentional use of speech pauses in wordspotting is reminiscent of the early days of automatic speech recognition (e.g. Rabiner, L.R.: "On creating reference templates for speaker-independent recognition of isolated words", IEEE Trans, vol. ASSP-26, no 1, pp. 34-42, February, 1978), where algorithmic limitations required the user to briefly pause between words. These early recognizers performed so-called isolated word-recognition that required the words to be spoken separated by pauses in order to facilitate the detection of word endpoints, i.e. the start and end of each word. One technique for detecting word endpoints is to compare the speech energy with some threshold value and identify the start of the word as the point at which the energy first exceeds the threshold value and the end as the point at which energy drops below the threshold value (e.g. Lamel, L.F. et al: "An Improved Endpoint Detector for Isolated Word Recognition", IEEE Trans., Vol.

ASSP-29, pp. 777-785, August, 1981). Once the endpoints are determined, only that part of the input that corresponds to speech is used during the pattern classification process. In this prior art technique, the pause is not analyzed and therefore is not used in the pattern classification process.

Speech recognition systems include those based on Artificial Neural Networks (ANN), Dynamic Time Warping (DTW), and Hidden Markov Models (HMM).

DTW is based on a non-probabilistic similarity measure, wherein a prestored template representing a command word is compared to incoming data. In this system, the start point and end point of the word is known and the Dynamic Time Warping algorithm calculates the optimal path through the prestored template to match the incoming speech.

The DTW is advantageous in that it generally has low computational and memory requirements and can be run on fairly inexpensive processors. One problem with the DTW is that the start point and the end point must be known in order to make a match to determine where the word starts and stops. The typical way of determining the start and stop points is to look for an energy threshold. The word must therefore be preceded and followed by a distinguishable, physical speech pause. In this manner, there is initially no energy before the word, then

the word is spoken, and then there is no energy after the word. By way of example, if a person were to say <pause> "one" <pause>, the DTW algorithm would recognize the word "one" if it were among the prestored templates. However, if the phrase "recognize the word one now" were spoken, the DTW would not recognize the word "one" because it is encapsulated by other speech. No defined start and end points are detected prior to the word "one" and therefore the speech recognition system can not make any determination about the features of that word because it is encapsulated in the entire phrase. Since it is possible that each word in that phrase has no defined start point and end point for detecting energy, the use of Dynamic Time Warping for continuous speech recognition task has substantial limitations.

In the Artificial Neural Network approach, a series of nodes are created with each node transforming the received data. It is empirical (probabilistic) technology where some end number of features is entered into the system from the start point and the outpoint becomes the probabilities that those features came from a certain word. One of the major drawbacks of ANN is that it is temporally variable. For example, if a word is said slower or faster than the prestored template, the system does not have the ability to normalize that data and compare it to the data of the stored template. In typical human speech, words are often modulated or vary temporarily, causing problems for speech recognition based on ANN.

The Artificial Neural Network is advantageous in that its architecture allows for a higher compression of templates and therefore requires less memory. Accordingly, it has the ability to compress and use less resources in terms of the necessary hardware than the Hidden Markov Model.

The Hidden Markov Model has several advantages over DTW and ANN for speech recognition systems. The HMM can normalize an incoming speech pattern with respect to time. If the templates have been generated at one cadence or tempo and the data comes in at another cadence or tempo, the HMM is able to respond very quickly. For example, the HMM can very quickly adjust for a speaker using two different tempos of the word "run" and "ruuuuuuun." Moreover, the HMM process data in frames of usually (16 to 30 milliseconds), allowing it to have a very fast response time. Since each frame is processed in real time, the latency for HMM is less than for DTW algorithms which require an entire segment of speech before processing can begin.

Another advantage which distinguishes the HMM over DTW and ANN is that it does not require a defined starting or end point in order to recognize a word.. The HMM uses qualitative means of comparing the features in an input stream to the stored templates eliminating the need to distinguish the start and end points. It uses a statistical method to match the sound that is being detected with any sound

that is contained in its templates and then outputs a score which is used to determine a match. Although the HMM is superior to its counterparts, it is known from the prior art that its implementation to commercial fixed-point embedded systems (which are clearly different from PC platforms) has been neglected.

Many prior art speech recognition systems have a detrimental feature with respect to command word template generation. When templates are generated from data produced by recorded human speech, they may not accurately represent the way every person says a command word. For example, if a user's particular speech pattern differs significantly from the template data, then very poor performance from the speech recognition system will be experienced when compared to a user whose speech pattern is more similar to the template data. In an HMM recognizer, words are scored by their probability of occurrence. The closer a word is to its prestored template; the higher its probability is calculated. In order for a word to be considered a match, a preset decision threshold is used. In order to be recognized, the similarity between the uttered word data and the template has to exceed the preset decision threshold. Many speech recognition systems have not provided the user with any means of adjusting the preset decision threshold.

SUMMARY OF THE INVENTION

The invention solves the above-identified problems of the prior art by requiring the user to pause at least in between individual words of an audio or voice command. As an example, the command to turn on the lights becomes "lights<pause>on" or "<pause>lights<pause>on<pause>." Viewing each pause of detectable duration as a substitute for one or more syllables, this new command exhibits the same number of syllables as "turn lights on" and "please turn the lights on", respectively. This improves the FOM without requiring the speaker to memorize more words and the required word order in a voice command.

Accordingly, it is important to note the following key differences between the use of speech pauses in the present invention and in the prior art isolated-word recognition:

1) The invention treats the speech pauses as part of the keywords and as such treats those just like any other speech sound. Thus, the particular spectral qualities of the input signal during speech pauses are essential for a keyword to be correctly detected. In contrast, the prior art isolated-word recognition discards speech pauses during a pre-processing step; and

2) The purpose of the speech pauses in the present invention is to make the keywords longer rather than to simplify endpoint detection. In fact, no explicit detection is performed at all in the present invention.

It is therefore an object of the present invention to provide a system and method for more accurately recognizing speech commands without increasing the number of individual command words.

It is a further object of the invention to provide an apparatus for controlling an electrical device, such as a lighting fixture including an incandescent lamp or any other suitable electrical load by speech commands.

It is an even further object of the invention to provide a means for adjusting the threshold comparison value between prestored speech recognition data and uttered audio data to thereby accommodate users with different voice patterns.

Other objects, advantages and features of the invention will become apparent upon reading the following detailed description and appended claims, and upon reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred embodiments of the present invention will hereinafter be described in conjunction with the accompanying drawings, wherein like designations denote like elements throughout the drawings, and wherein:

FIG. 1 shows a block diagram of an apparatus for voice-activated control of an electrical device according to the present invention;

FIG. 2 is an output timing diagram showing the AC output delivered to a lighting fixture or other similar load;

FIG. 3 is a process flow chart for use with the invention for recognizing the presence of a voice command;

FIG. 4 is a process flow chart according to the invention for recognizing the presence of a voice command;

FIG. 5 is a chart schematically representing an acceptable energy level of voice command to background noise for actuating an electrical device;

FIG. 6 is a chart schematically representing an unacceptable energy level of voice command to background noise for actuating an electrical device;

FIG. 7 is a process flow chart according to a further embodiment of the invention for recognizing the presence of a voice command; and

FIG. 8 is a process flow chart according to an even further embodiment of the invention for recognizing the presence of a voice command.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring now to the drawings, and to FIG. 1 in particular, a functional block diagram of an apparatus **10** for receiving and processing voice commands according to the present invention is illustrated. The apparatus **10** comprises an audio input **12** that is monitored by a micro-controller **14** which performs processing functions associated with a programmed speech recognition algorithm in an attempt to identify voice commands. Once a voice command has been identified, the micro-controller **14** directs a dimmer control circuit **16** or other electrical control or switching circuit to take an appropriate action. The power required for operation of the micro-controller and other associated electronic units comes from a conventional 5 V DC power supply **18**, that may be converted from an alternating current voltage input **20**. Once a voice command has been recognized by the micro-controller **14**, the control circuit **16** is manipulated to

provide appropriate AC (or DC) output **22** to an external electrical device, such as a lamp, electrical outlet, and so on.

The micro-controller **14** preferably includes an 8-bit or 16-bit MCU (micro controller unit) embedded, fixed-point microprocessor as the central processing unit (CPU), an on-chip analog to digital (A/D) converter, a Read Only Memory (ROM) bank, a Static Random Access Memory (SRAM) bank, and general purpose Input-Output (I/O) ports. This embedded system is specifically different from a PC system, in that it does not have the ability to be modified once the program has been loaded into the chip; it does not have a disk drive or any other means of being changed. It has firmware as opposed to software, and it does not have a monitor or a keypad. It has a very specific purpose that once the system is loaded with the firmware (hardware), it is only allowed to perform one specific task which is to analyze incoming audio data and comparing that data to one or more sets of prestored data to thereby control an outside electrical device. Unlike a PC which has a general purpose processing unit that can do a multitude of tasks from word-processing to game play to speech recognition, for instance, the system of the present invention that runs the Hidden Markov Model is embedded and all contained on a single PC board. The microphone and all of the analog and digital circuitry, including all of the control circuitry and means for interacting with the speech recognition system is preferably contained on a single printed circuit board. The printed circuit board with all of its components is preferably

enclosed in a stand-alone housing (not shown), such as a plastic box, that has an electrical plug for connection with an electrical receptacle. Preferably, the housing also includes electrical receptacles for accepting electrical devices of one type or another to be controlled by voice command.

The audio input **12** may be in the form of a microphone or similar device for inputting audio signals into the apparatus **10**. The input signals are then filtered and sent to two amplifiers (not shown) having different gains. The micro-controller processes the analog signal from the channel with the highest gain unless there are data over-ranges. When over-ranges are encountered, the channel with lower gain is utilized. This effectively provides for more dynamic range in the input audio signal without the delay associated with other Automatic Gain Circuits (which have a time delay associated with correction of gain).

An audio output device **24** may be provided as feedback means and instructions to the user. Compressed speech may be stored in System Memory and played back to the user at the appropriate time. This synthesized speech can be used for prompting during a training process where the user(s) is prompted to speak for creating voice templates prior to use. The Audio Output device **24** is preferably in the form of a speaker element.

The apparatus **10** interfaces directly with any source of AC power **20** in general and is specifically adapted to interface with a common household AC (120Vrms in the USA) power outlet via a connector plug (not shown). The connector plug is polarized, with the larger of the two prongs connecting to "common" (white) of a normal household power outlet. The "hot" (black) input is fused with a 5 amp fuse as a safety measure.

The AC Input **20** is connected to the Dimmer Control Circuit **16** as well as to a full-wave rectifier which is part of the DC Power Supply **18**.

The Dimmer Control Circuit **16** controls the current delivered to a lighting fixture (not shown) or other similar load and is regulated by a TRIAC. In this manner, it is possible to turn the fixture on or off, so it can be dimmed without dissipative power losses. The TRIAC is driven by an opto-coupler and is therefore electrically isolated from the circuitry of the apparatus **10** which contains the audio input amplifier, the micro-controller **14**, and other related functional circuitry.

The input AC signal feeds into a zero-crossing detection circuit (not shown) and generates a trigger pulse which is used as the clock input for several one-shots with time constants adjusted to provide pulse trains at 33% and 66% of the AC line frequency duty cycle. These pulse trains are supplied to a multiplexer (not

shown) which is controlled by the micro-controller 14 adapted for selecting an appropriate digital pulse train to drive the opto-coupler.

Several bits from Port O of the speech recognition chip (bits 2 and 3) select the appropriate channel of the multiplexer which is adapted for driving this circuit. It has four states:

BIT 1	BIT 0	STATE
0	0	ON
1	0	2/3 ON (dim)
0	1	1/3 ON (low dim)
1	1	OFF

For the 1/3 ON dim selection, the TRIAC is on 33% of the time, allowing current to flow to the lighting fixture or other similar load. When the TRIAC turns off, the current does not flow to the fixture. This causes the lighting fixture to appear dim. In this technique, power is not dissipated by the dimmer circuit when the lights are dimmed. Output waveforms for each of the above states are illustrated in FIG. 2.